

情报理论与实践  
*Information Studies: Theory & Application*  
ISSN 1000-7490, CN 11-1762/G3

## 《情报理论与实践》网络首发论文

题目：疫情防控中的大数据应用伦理问题研究  
作者：丁波涛  
网络首发日期：2020-12-21  
引用格式：丁波涛. 疫情防控中的大数据应用伦理问题研究. 情报理论与实践.  
<https://kns.cnki.net/kcms/detail/11.1762.G3.20201221.1308.002.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

●丁波涛（上海社会科学院信息研究所，上海 200235）

## 疫情防控中的大数据应用伦理问题研究\*

**摘要：**采用大数据技术提升疫情防控能力，已成为世界各国的共识，但这也会引发许多新的伦理挑战。文章对大数据应用于疫情防控的伦理风险点进行了梳理，并从个体利益与公共利益、风险与收益、效率与公平平衡的角度，分析疫情数据伦理问题的特殊性，提出规范数据获取、落实算法理性和限制成果传播等监管建议。

**关键词：**大数据；疫情防控；数据伦理；信息安全

### A Study on Ethics of Big Data Applied to Epidemic Prevention and Control

**Abstract:** The adoption of big data to improve the ability of epidemic prevention and control has become a consensus of the world, which will also, however, cause many ethical challenges. The paper sorted out the ethical risks in application of big data to infectious diseases prevention and control, and the solution of ethical issues were analyzed from the perspective of the balance of individual and public interests, risks and benefits, efficiency and fairness. In addition, the paper proposed the countermeasures and suggestions such as standardizing data access, implementing algorithmic rationality and limiting the dissemination of results.

**Keywords:** big data; epidemic prevention and control; data ethics; information security

## 1 背景与概念

### 1.1 问题的提出

近 20 年来全球爆发多次严重疫情，给各国造成巨大损失，促使人们开始对疾病监测与预防体系进行反思。传统疾控系统局限于分析诊断、医学检测以及受控统计等医疗数据，反应较慢而且难以及时发现疫情产生和传播的社会关联，已经不能适应新型传染病防控工作的需要<sup>[1]</sup>。在数字化的时代背景下，采用大数据、云计算、人工智能等数字技术提升疫情防控能力，已成为世界各国的共识。

2002 年，Eysenbach 提出了“信息流行病学”（Infodemiology）的概念<sup>[2]</sup>，并逐步演化为“数字流行病学”（Digital Epidemiology）。无论其名称如何变化，其核心都是将医疗数据与非医疗数据（社交媒体、个人行为痕迹、购物清单、移动数据等）结合起来，并利用数字技术等进行深入分析，从而更早地察觉疫情，并更好地预测传播模式和传播路径，为公共卫生干预提供支撑。

流行病学需要研究人类宿主、病原体和中介之间复杂相互作用<sup>[3]</sup>。传统流行病学研究面临的一大障碍是用于分析人类活动模式的时空数据十分有限。而互联网日益普及、移动设备爆炸性增长、在线平台快速发展，其生成的海量数据应用于公共卫生，可为疫情防控提供更广域的视野和更敏捷的洞察力：一是线下与线上数据相结合，监测更全面；二是在疾病爆发之前发现风险，预警更及时；三是将疾病发展与社会背景相关联，判断更准确；四是将本地数据与全球数据相融合，防控更广泛。

然而大数据技术是“致广大”与“致精微”的内在贯通<sup>[4]</sup>，其应用往往是双刃剑：一方面有助于破解传统疫情防控中存在的信息来源单一、处理速度慢、分析精度差等难题，推动疫情防控转向高效、协同、精准和智能；另一方面，多源大数据的融合、医疗与非医疗大数据的关联、医疗机构和数字企业的跨界，也会引发或放大许多伦理挑战，这是不容回避的重要课题。

### 1.2 数据伦理的概念界定

“数据伦理”一词则来自于“信息伦理”。1988 年罗伯特·豪普特曼（R. Hauptman）

\* 本文为上海市 2020 年度“科技创新行动计划”软科学重点项目“数字流行病学中的信息安全风险与防范对策研究”的成果，项目编号：20692106300。

首次提出“信息伦理”一词，他认为“所有对与信息生产、信息储存、信息访问和信息发布伦理问题相关的研究都统称为信息伦理”<sup>[5]</sup>。我国学者吕耀怀则认为信息伦理是指“涉及信息开发、信息传播、信息的管理和利用等方面的伦理要求、伦理准则、伦理规约，以及在此基础上形成的新型的伦理关系”<sup>[6]</sup>。

目前关于数据伦理的定义较多，如“数据(包括生成、记录、管理、加工、传播、共享和使用), 算法(包括人工智能、智能代理、机器学习和机器人)和相关实践(包括创新、编程、黑客和专业编码)中的伦理相关问题”<sup>[7]</sup>。无论何种定义，数据伦理都有三个特征：一是主要关注个人数据，一般不包含媒体、新闻、图书馆和情报工作中的信息伦理；二是覆盖数据价值链的全部环节，包括创造、采集、管理、共享、分析、传播、利用等；三是目标是减少偏见歧视、建立各方信任、实现公平正义、促进积极向善，使大数据更好造福人类。而从数据伦理所涉及的范围来看，可分为三个层面，一是个体层面，即数据开发利用对民众个人权益的影响；二是社会层面，即对个体与个体之间关系的影响；三是国家层面，即对国与国关系的影响。

## 2 研究现状

医疗健康数据应用的伦理风险受到广泛关注，是当前国内外研究热点。按研究的角度与层面，可将现有相关成果分为四类：一是个体信息权利层面的研究，着重分析医疗大数据应用与公民信息权利之间的冲突。如 L. Gwendolyn 等<sup>[8]</sup>指出知情的公众讨论、更大的透明度和新的伦理框架对于建立公众对使用新技术控制传染病的信任至关重要；刘星<sup>[9]</sup>、B. M. Knoppers 等<sup>[10]</sup>分析了信息安全与隐私、知情同意、风险与收益；郭向凝<sup>[11]</sup>、滕长利<sup>[12]</sup>分析了医疗大数据采集以及开发与患者隐私权冲突；包桢冰<sup>[13]</sup>、王强芬<sup>[14]</sup>等分析了医疗大数据中的自主性、主体性、被遗忘权等问题。二是相关利益方关系层面的研究，着重分析医疗大数据应用引发的医疗机构、大数据企业以及医生、患者、民众之间伦理关系调整。如 M. Mirchev<sup>[15]</sup>讨论了医疗数据开发中个人利益与公共利益的冲突；刘辉<sup>[16]</sup>、蔡雷<sup>[17]</sup>、C. Brall 等<sup>[18]</sup>讨论了医学大数据应用中的利益分配不公平以及数字医疗鸿沟；李晓洁等<sup>[19]</sup>以 Google 流感预测为例分析了“大数据的傲慢”以及被动共享、算法黑箱；Eric Topol<sup>[20]</sup>、L. F. Hogle<sup>[21]</sup>讨论了数字医疗引发的医生地位、医患关系、医学研究者与医疗工作者之间界限的变化。三是法规制度层面的研究，着重探讨如何解决医疗大数据伦理问题。王丽莎研究了医疗大数据应用的法规和伦理规制<sup>[22]</sup>；刘瑞爽<sup>[23]</sup>分析了 GDPR 对我国医学研究伦理审查的启示；张海洪<sup>[24]</sup>分析了健康数据二次使用研究的伦理审查制度；D. Golinelli 等<sup>[25]</sup>讨论了国家卫生部门授权第三方机构使用国民医疗健康数据中的伦理与法律规范；B. D. Mittelstadt 等<sup>[26]</sup>提出了 11 个关键主题，以指导医学大数据伦理审查和治理。四是由国内外机构提出的医疗大数据伦理框架。如北京大学健康医疗大数据国家研究院牵头起草了《关于健康医疗大数据优良实践的伦理共识》(第一版)，提出了医疗大数据应用中的各方权责<sup>[27]</sup>；新加坡科学、卫生和政策伦理协会(SHAPES)发布的《健康与研究大数据伦理框架》，提出了医疗大数据应用的 9 个基础价值观和 7 个过程价值观<sup>[28]</sup>。

上述研究为疫情防控中的大数据应用伦理研究提供了坚实基础，但仍有待深入之处：其一，现有研究多立足于大数据在疾病诊治、医学研究等个体场景中的应用，而疫情防控中的大数据应用属于公域场景，其社会关联更加复杂；其二，现有研究所涉及的大数据多限于医疗卫生领域的的数据，而疫情防控中所使用的大数据范围更广。正因为如此，现有研究中提出一些大数据伦理基本原则，如知情同意、匿名化、最小使用等，在疫情防控领域未必完全适用。疫情防控大数据应用中的一些新问题，如可能引发的社会恐慌、群体歧视等，在现有研究中也很少涉及。因此，本文将紧密结合疫情大数据应用的特点，深入讨论其中的数据伦理问题，并提出应对思路。

## 3 大数据应用于疫情防控中的伦理难题

根据国际数据管理协会(DAMA)的定义，数据生命周期包括数据规划、规范定义、创建或获取、存储、维护、分析、处理、传播、使用、清除等<sup>[29]</sup>。为简便起见，本文将数据应用全过程分为数据获取、分析、传播三个环节，并从数据伦理产生的个体、社会和国家三个层面，梳理疫情大数据应用中的伦理问题，如表 1 所示。



表 1 疫情防控中的数据伦理问题

	个体层面	社会层面	国家层面
数据获取	✓ 对个人知情同意权的影响 ✓ 对数据匿名化的影响	✓ 数据干扰与扭曲	✓ 数字鸿沟与全球正义
数据分析	✓ 算法可靠性	✓ 算法黑箱 ✓ 群体性歧视	
数据传播		✓ 怀疑泛化 ✓ 非授权发布（机构）	✓ 非授权发布（国家）

### 3.1 数据获取中的伦理问题

1) 个人知情同意。知情同意是患者的最基本权力，是当代医学伦理的核心要求，被越来越多国家的司法所承认和保护<sup>[30]</sup>。然而传染病大规模爆发的场景中，要求在采集患者和潜在感染者数据之前要征得每个人的明确同意，无论是在时间上或是在精力上都不可行；同时在疾病传播中，即使大部分人同意，但只要少数人不同意，就会造成数据链路破坏，无法开展流行病学分析。同时患者的知情同意权行使还面临一个难题：采集者很难知晓所收集数据的最终用途。疫情数据开发通常会经过一个价值链：从数据收集到聚合、分析、决策和执行，其中任何环节都可能出现道德问题，最终执行者使用数据的目的可能完全不同于数据采集者的最初意图，从而造成数据使用范围超出采集者当初与患者的约定<sup>[31]</sup>。

2) 数据匿名化。为减少潜在风险，个人数据往往采用匿名化方式进行保护。然而疫情数据分析通常需要关联多源数据，如将医疗数据与患者行踪关联起来。如果不能识别数据中的个人身份，就无法实现这种关联，达不到疫情防控的目标。即使这些数据不是实名的，也至少需要能够识别其唯一身份的假名标识<sup>[32]</sup>。同时数据分析技术飞速进步，当不同来源的数据进行关联后，即使是数据匿名化也无法充分保护数据源的身份，例如许多实验发现，在一个机构中借助于年龄、籍贯、性别等少数几个数据就足以识别出大部分人的身份<sup>[33]</sup>。在这种情况下，所谓的数据匿名化也失去了意义。

3) 数据干扰与扭曲。大数据提升了疫情监测能力，但也会导致监测体系对数字环境高度敏感，一些人可能会有意或无意发布一些虚假的疾病信息，从而造成系统错误判断。对于故意在网上发布虚假信息的行为，可以通过传染病防治、互联网信息管理等法规进行约束和惩处，但其他一些间接发布“错误”信息的行为却难以用法律来规范。例如，许多监测平台都是根据民众网上行为与疾病的相关性来判断当地是否出现疫情。但如果某地并未发生疫情却有大量网民在网上搜索、浏览和传播与疾病相关的内容，也会向系统发出错误信号，导致错误预警。当年 Google 的“流感趋势”项目之所以失败，就是因为非疫区民众大量搜索和浏览有关流感的内容引起“信号扭曲”<sup>[34]</sup>。

4) 数字鸿沟与全球正义。与传统方法相比，通过现代信息技术获取医疗卫生数据的成本更低。尤其是一些落后国家的流行病监测体系还不完善，但其互联网和移动设备却相对普及，大量民众在网上浏览和发布信息，借助互联网获得流行病数据比传统医学渠道更为可行。然而，对互联网进行大规模监测需要强大的技术和资金支持，通常只有发达国家具备此能力，而且几乎所有重要的全球性网络平台（如 Google、Facebook、Twitter 等）都由美国掌控，其在获取公共数据方面具有天然优势，甚至在获取他国公民网上数据时，比该国政府更有优势。由此发达国家可以更早发现和应对疫情，从而进一步拉大南北国家在公共卫生上的差距。发达国家甚至可能将之作为国际政治武器，利用其在疫情数据获取和监测中的优势地位对其他国家施加影响。

### 3.2 数据分析中的伦理问题

1) 算法可靠性。错误的疫情预测结果可能会对相关地区的经济社会造成巨大破坏。例如 2007 年 Google 曾错误地预测非洲某国将出现霍乱，引发该国社会恐慌，并挤占了其本就十分匮乏的医疗资源<sup>[35]</sup>。因此，疫情防控中所使用的大数据算法必须高度稳健和可靠。然而大数据分析只能揭示事物之间的相关关系，而非因果关系<sup>[36]</sup>，因此大数据预测不可能百分之百准确，而必会出现“假阳性”结果。从科学角度上讲，一定比例的预测误差是无法避免的；但从社会角度而言，任何失误都可能引发社会动荡和民众不满。

2) 算法黑箱。新一代人工智能所依靠的深度学习算法，在输入数据和输出结果之间存

在着人类无法洞悉的“隐层”，可称之为“算法黑箱”<sup>[37]</sup>。黑箱的存在，一方面导致算法运作不可解释、不可监察，使人类难以真正理解和有效控制算法，从而无法预见和应对传染病分析算法可能带来的风险；另一方面可能导致疾控部门过度依赖分析算法，只关注大数据分析结果而不深究其背后的流行病学机理，一旦出现算法失灵，将造成疫情监控的巨大漏洞。

3) 群体性歧视。传染性疾病往往在一些相互密切接触的群体中传播，而大数据算法会“忠实地”揭示出疫情与群体的相关性，按照职业、区域、种族、收入等维度将人群划分成不同的风险等级。这种风险分类无疑具有科学意义上的正确性，但向社会公布之后很可能会造成对特定群体的歧视。许多传染病爆发时都会出现群体歧视现象，但毕竟发生在疾病确诊之后。而基于大数据的流行病学分析只是在疾病确诊前做出的推测，其结果只能表明一定可能性和潜在风险，由此造成的群体歧视范围更广、伤害更大，也更加背离了平等包容的基本社会准则。

### 3.3 结果传播中的伦理问题

1) 怀疑泛化。当基于大数据的流行病学分析发现某些人可能被病毒感染时，出于保护公共卫生安全的需要，政府部门将不得不采取措施，对这些人群进行监视并限制其活动。但问题在于，仅仅根据大数据分析结果便对特定人群采取监视和隔离措施，在法律和伦理上是否合理？现代法律的一项基本原则是“无罪推定”，即任何人在被证明有罪之前都是无罪的。同理，一个人在被确诊或被证实为密切接触者之前，都应当被认为是健康的，否则将产生所谓“怀疑泛化”（The generalization of suspicion）问题<sup>[38]</sup>。如果不能把握好分析结果的尺度而出现“怀疑泛化”，将会引起社会恐慌、人人自危，也会增加传染病控制的难度，导致医疗和社会资源浪费。

2) 非授权发布。一些互联网和大数据企业等非医疗卫生机构，其分析和预测结果可不经传统医疗部门的审查，而直接在互联网上进行发布。而大众通常不具备正确理解这些数据和结论的专业知识，难以区分大数据预测和临床确诊的不同含义，由此产生的分析结果误读可能会引发不必要的社会恐慌<sup>[39]</sup>。虽然许多国家严格限定了疫情发布权限，但一些机构仍可以科研成果、疾病预测等名义发布关于传染病流行的分析结果，或者是在互联网上公开发布其他国家的流行病预测结果（比如说在美国发布中国的流行病预测），从而避开法规对疫情发布权限的限制。

## 4 疫情大数据治理中的三个平衡

从上述分析中可以看出，在疫情防控的场景中，大数据治理面临着许多新的挑战。然而大数据所具有的数据主体、数据来源和数据价值多元性<sup>[40]</sup>，决定了大数据治理往往并无最优解，而是需要兼顾个人权利与公共善之间、目的善与手段善、价值理性与技术理性的平衡，使大数据利用行为得到伦理确证与道德规约<sup>[41]</sup>。

### 4.1 个人权益与公共利益的平衡

从法律上讲，在个体利益与公共利益冲突的场景下，信息保护与信息利用应遵循比例原则，统筹考虑个人信息的敏感性和公共需要的紧迫性之间的平衡<sup>[42]</sup>。个人数据无疑关系到个人隐私，但它也具有公共卫生功能，利用这些数据改善人类健康，所有人都将从中受益。而且随着我国经济社会发展，人口流动的规模与速度都不断增加，不依靠大数据已难以对重大突发事件进行有效处置。因此在面对重大疫情时，为更好保护公众利益，我国应将个人数据利用置于更优先位置，而个人信息权利则应有所克减。当然，大数据开发优先并非不保障个人隐私，而是采取“积极保障”策略，即在开发数据的同时加强隐私保护，而不是通过封闭数据来保障隐私的“消极保障”。

### 4.2 数据收益与技术风险的平衡

虽然数据安全风险防范十分重要，但是零风险必然意味着零收益，关键是在风险和收益之间保持适当的平衡<sup>[43]</sup>。伦理学角度的义务论也认为大数据应用应坚持“效用原则”，即评估一个行动或决策可能引起的受益与风险比是否合适、是否可接受<sup>[44]</sup>。大数据为流行病监测与预警提供了新的手段与方法，有利于保障民众生命安全和和谐稳定。如因厌恶风险而放弃大数据应用，对全社会而言将是巨大损失。因此，在疫情大数据伦理问题上应坚持“数据主义”与“人本主义”相结合的原则，即一方面坚持数据优先、普遍互联<sup>[45]</sup>，积极利用大数据来防控重大传染病疫情；另一方面，大数据利用也需要人本主义情怀，从事数字流行病学研究的机构在数据采集、数据分析和成果发布时应采取极为慎重的态度。

4.3 技术效率与社会公平的平衡

疫情防控需要海量数据资源与复杂数字技术相结合，可能会造成多个层面的不平等。一是个体层面上，潜在疫区的民众在医学确诊前就可能被限制甚至隔离，而区域之外的民众获得了更充足的应对时间；二是机构层面上，掌握数据并具备强大分析能力的机构，将与普通民众之间形成巨大的信息和权利不对称；三是国家层面上，数字技术发达的国家将在传染病防治方面获得更大的优势，进一步拉大国家间差距。

数字流行病学涉及民众的生命安全、社会的公平正义以及国家的主权完整，具有极大特殊性，因此疫情防控中的大数据应用应当以“效率与公平并重”为原则，通过技术透明化、限制使用、限制传播等方式，保障大数据更好地服务于公共利益，同时减少对某些群体产生的不利影响，从而弥合数字技术使用造成的不平等。

5 加强疫情大数据开发利用中的伦理监管

综上所述，大数据将大幅提升疫情防控能力，但也蕴藏着巨大伦理风险。对此，人们既要包容的心态促进新技术应用，也要用审慎维护疫情大数据开发利用中的三个平衡，通过规范数据获取、落实算法理性和限制结果传播等监管举措，来防范数据伦理风险，如图 1 所示。

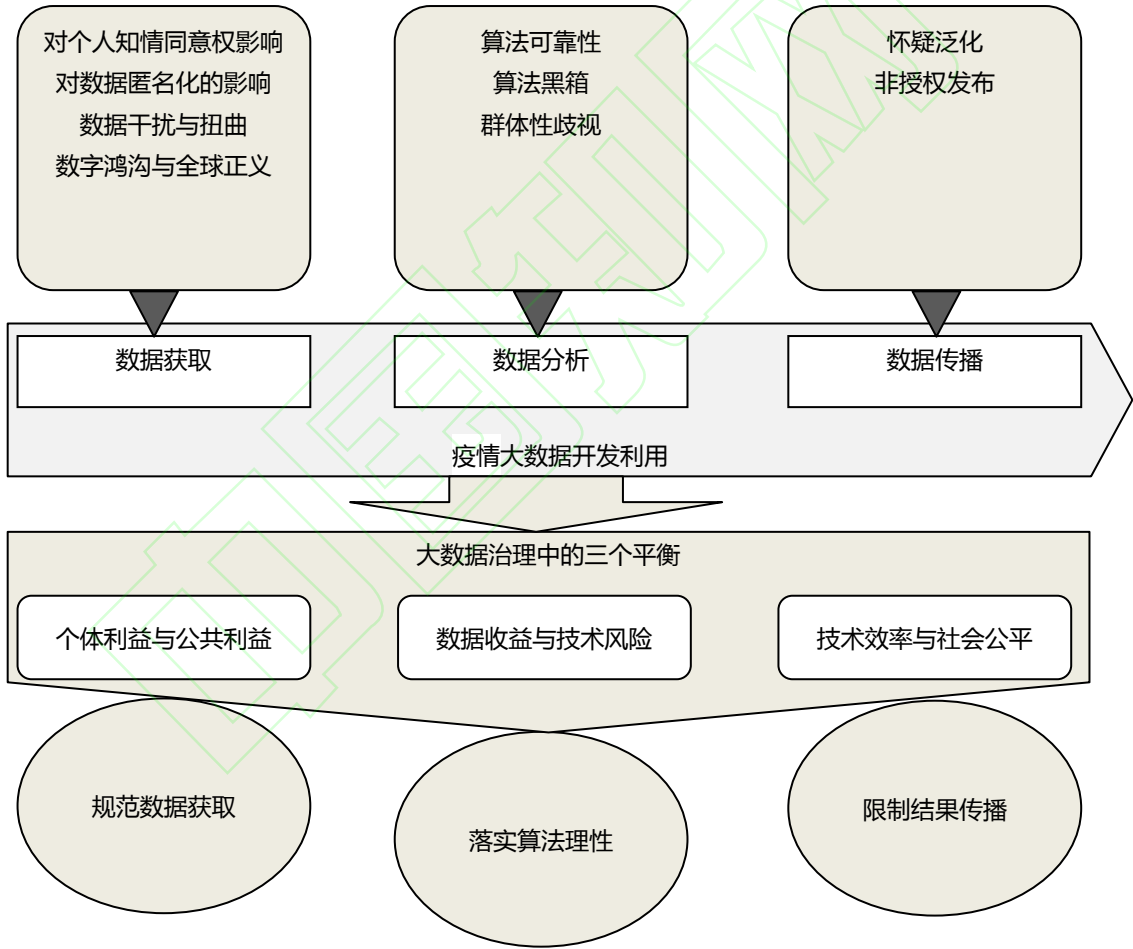


图 1 疫情大数据伦理监管的主要对策

5.1 规范数据获取

在面对重大疫情时，对于疫情防控中所需要的个人数据，建立原则上可不经个体同意、允许直接收集和使用的规则，转而将重心放在监管、规制如何合理使用这些健康医疗数据之上<sup>[46]</sup>。

由于疫情防控中的大数据使用既不宜匿名而且难以匿名，为降低风险，各类机构获取数据的行为应当受到一定限制。一是主体限制。只有政府授权的疫情防控和公共卫生部门可不



经个体同意、直接收集和使用个人数据，其他机构开展流行病学分析和预测以其合法创造和采集的数据资源为限。二是用途限制。各类机构只能因特定用途去收集与使用数据，其获取的可直接识别主体的个人医疗与健康数据只能服务于传染病监测与防治等公共目的，不得用于与疫情防控无关的其他事务，更不得与其他机构共享或对外泄露。三是期限限制。疫情结束后，相关机构除保留不涉及个人私密的过程和结果信息之外，对所有能识别出个人身份的原始信息应当全部删除。例如 Google 在进行流行病分析时，通过“在特定时间内删除用户身份信息”等方式保护公民隐私。

## 5.2 落实算法理性

一是算法透明。透明度可以帮助防止滥用机构权力，同时也鼓励个人安全地分享更多的相关数据，为社会做出更好的大数据预测<sup>[47]</sup>。虽然目前人们对算法透明的必要性、有效性、可行性等存在许多争论<sup>[48]</sup>，但一般而言，对具有垄断地位的算法，或国家财政资金提供支持的、目的是提供普遍公共服务的算法，人们应有权要求其公开<sup>[49]</sup>，服务于疫情防控的大数据算法显然属于此类。当然考虑到算法的专业性、机密性和版权等因素，算法透明不等于算法源代码公开，而是公开算法的假设、逻辑、模型、功能、风险等。

二是算法问责。算法透明是基于本质主义的事前规制，而算法问责则是基于实用主义的事后规制，实际中后者相比前者往往更具普适性和规制力<sup>[50]</sup>。算法问责意味着利用大数据开展流行病学分析的机构应对其所使用算法引发的危害负责，无论这些机构是否能解释算法是如何产生结果的，即这些机构可能需要承担无过错责任。由于绝大多数算法都具有超出一般民众理解力的复杂性，算法透明并不能形成对数据控制者的有效监督，而算法问责制可以向数据控制者施加强大压力，促使其开发能确保不会造成伤害的最佳算法<sup>[51]</sup>。

三是算法和医学协作。大数据技术并不能替代专业医学方法，而是应当相互配合，将大数据更快而专业医学更准的优势结合起来。大数据的作用是发现疫情苗头和线索，但只有在进行专业医学检测确认后，才能公开发布疫情信息。例如，西方国家最先报道新冠肺炎疫情的机构是加拿大蓝点公司，其监测平台可以扫描和分析 65 种语言的外国新闻报道、航空公司数据和动物疾病爆发报告等信息，平台发现可疑信息之后，先由流行病学家对其进行审核、判断并确诊后再向外发布<sup>[52]</sup>。再如针对 Google 流感平台（GFT）预测不准的问题，有学者建议将美国疾控中心（CDC）数据与 GFT 分析结果相结合，经过“重新校准”的信息将更加可靠<sup>[53]</sup>。

## 5.3 控制成果传播

我国《传染病防治法》以及《互联网信息服务管理办法》、《网络安全法》等规定，任何机构不得发布扰乱经济秩序和社会秩序的虚假信息，不得侵犯他人名誉、隐私、知识产权和其他合法权益。然而传染病疫情属于高度敏感的信息，相关机构对外发布数字流行病学分析结果，即使是真实的，也可能引起民众恐慌、社会动荡，或引起对特定群体的歧视。

因此对数字流行病学分析结果的传播应当提出更高要求，不仅必须是“真实的”，还必须是“无害的”，即不应伤害他人或特定群体的隐私权、名誉权和正常生产生活的权力。这要求从事流行病学研究的成果加强对研究成果的伦理审查，并采取“内外有别”的成果传播方式：对那些满足真实性、合法性和无害性原则的成果可以直接对外公开发布；而对那些“真实但可能有害”的成果，则只能提供给专门政府部门用于疫情扩控，而不应直接对外公布。关于成果传播控制的实现方式，在国内可以依靠立法和行业自律，国家之间则需要通过协商来形成共识。

# 6 结论与展望

大数据既是解决问题的重要手段，但也会带来许多新的安全挑战<sup>[54]</sup>。虽然目前大数据在传染病防控中的应用尚处于起步阶段，但未雨绸缪地及早研究其中的伦理问题、预先规避风险是十分必要的。本文对大数据应用于传染病防控中的伦理风险进行了梳理，并从个体利益与公共利益、风险与收益、效率与公平平衡的角度，分析疫情防控中数据伦理问题的特殊性，提出了规范数据获取、落实算法理性和限制成果传播等对策建议。

总体而言，当前国内外对传染病防控中大数据应用的研究与实践还不充分，许多伦理问题只有在深入应用的过程中才会逐步显现。未来人们一方面要对伦理问题保持高度关注，对大数据应用于传染病防控中的各种伦理风险及早察觉和研究；另一方面也需要社会各界共同努力，通过政策与技术、社会监督与行业自律相结合，以充分规避大数据应用中的伦理风险，

实现科技向善的根本目标。□

#### 参考文献

- [1]ECKMANN S T, FÜLLER H, ROBERTS S L. Digital epidemiology and global health security; an interdisciplinary conversation[J]. Life Science, Society and Policy, 2019, 15(2):1-13.
- [2]EYSEN BACH G. Infodemiology: the epidemiology of (mis)information[J]. American Journal of Medicine, 2002, 113(9):763-765.
- [3]GUTIERREZ J B, GALINSKI M R, CANTRELL S, et al. From within host dynamics to the epidemiology of infectious disease: scientific overview and challenges[J]. Mathematical Biosciences, 2015, 270(12): 143-155.
- [4]田海平. 生命医学伦理学如何应对大数据健康革命[J]. 河北学刊, 2018(4):33-41.
- [5]杨晶晶, 谷立红, 田红. 信息伦理研究综述[J]. 电子政务, 2011(7):61-67.
- [6]吕耀怀. 数字化生存的道德空间——信息伦理学的理论与实践[M]. 北京:中国人民大学出版社, 2018:2.
- [7]FLORIDI L, TADDEO M. What is data ethics? [M].Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016:374-380.
- [8]GILBERT G L, DEGELING C, JOHNSON J. Communicable disease surveillance ethics in the age of big data and new technology [J]. Asian Bioethics Review, 2019(11):173-187.
- [9]刘星, 王晓敏. 医疗大数据建设中的伦理问题[J]. 伦理学研究, 2015(6):119-122.
- [10]MARIAKNOPERS B, MARKTHOROGOOD A. Ethics and big data in health[J]. Current Opinion in Systems Biology, 2017(3):53-57.
- [11]郭向凝, 张洪江. 医疗大数据的隐私伦理问题研究[J]. 锦州医科大学学报(社会科学版), 2019(3):21-24.
- [12]滕长利. 医疗大数据信息采集权与患者隐私权的冲突研究[J]. 卫生经济研究, 2019(5):21-23.
- [13]包桢冰, 徐佩. 医疗人工智能的伦理风险及应对策略[J]. 医学与哲学, 2018(6A):37-40.
- [14]王强芬. 大数据时代背景下医疗隐私保护的伦理困境及实现途径[J]. 中国医学伦理学, 2016(4):685-689.
- [15]MIRCHEV M. Patient information ownership in the age of digital health and big data[J]. European Journal of Public Health, 2019(11):412-421.
- [16]刘辉, 丛亚丽. 临床医学大数据的伦理问题初探[J]. 医学与哲学, 2016(10A):32-36.
- [17]蔡雷, 李伟桓, 师咏勇, 等. 医学大数据使用安全及伦理问题探讨[J]. 中国医学伦理学, 2019(8):1004-1009.
- [18]BRALL C, SCHRÖDER-BÄCK P, MAECKELBERGHE E. Ethical aspects of digital health from a justice point of view[J]. European Journal of Public Health, 2019(10):18-22.
- [19]李晓洁, 丛亚丽. 从“谷歌流感趋势”预测谈健康医疗大数据伦理[J]. 医学与哲学, 2019(14):5-8.
- [20]ERIC TOPOL M D. The creative destruction of medicine: how the digital revolution will create better health care[M]. Basic Books, New York, 2013: 90.
- [21]HOGLE L F. The ethics and politics of infrastructures: creating the conditions of possibility for big data in medicine[M]. MITTELSTADT B, FLORIDI L. The ethics of biomedical big data. Law, Governance and Technology Series(Vol 29). Springer, Berlin, 2016:125.
- [22]王丽莎. 互联网医疗大数据的法律与伦理规制研究[J]. 中国医学伦理学, 2017(11):1322-1325, 1342.
- [23]刘瑞爽. GDPR 对我国医学研究伦理审查的启示[J]. 医学与哲学, 2019(3):29-33.
- [24]张海洪. 健康数据二次使用研究的伦理审查探讨[J]. 医学与哲学, 2019(3):26-28.
- [25]GOLINELLI D, TOSCANO F, BUCCI A, et al. Transferring health big data between public and private companies within the European legal framework[J]. European



Journal of Public Health, 2018(11):225-226.

[26] MITTELSTADT B D, FLORIDI L. The ethics of big data: current and foreseeable issues in biomedical contexts[J]. Science and Engineering Ethics, 2016(22):303-341.

[27]刘瑞爽,冯瑶,李晓洁,等. 关于健康医疗大数据优良实践的伦理共识 [J]. 中国医学伦理学, 2020(1):8-11.

[28] XAFIS V., SCHAEFER G O, LABUDE M K, et al. An ethics framework for big data in health and research[J]. Asian Bioethics Review, 2019(11):227-254.

[29] DAMA International. DAMA 数据管理知识体系指南[M]. 马欢,刘晨,等,译. 北京:清华大学出版社, 2012:9.

[30]何卫涛. 论我国患者知情同意权制度的完善[D]. 厦门:厦门大学, 2011:10.

[31] GARATTINI C, RAFFLE J, AISYAH D N, et al. Big data analytics, infectious diseases and associated ethical impacts[J]. Philosophy Technology, 2019, 32(1):69-85.

[32] MITTELSTADT B, BENZLER J, ENGELMANN L, et al. Is there a duty to participate in digital epidemiology? [J]. Life Science, Society and Policy, 2018, 14(9):1-24.

[33] WHITE S E. De-identification and the sharing of big data[J]. Journal of Ahima, 2013, 84(4):44-47.

[34] LAZER D, KENNEDY R, KING G, VESPIGNANI A. The parable of Google Flu: traps in big data analysis[J]. Science, 2014, 343(14):1203-1205.

[35] ECKMANN S, FÜLLER H, ROBERTS S L. Digital epidemiology and global health security: an interdisciplinary conversation[J]. Life Science, Society and Policy, 2019, 15(2):1-13.

[36]王天思. 大数据中的因果关系及其哲学内涵[J]. 中国社会科学, 2016(5):22-42.

[37]许可. 人工智能的算法黑箱与数据正义[N]. 社会科学报, 2018-03-29 (6) .

[38] SAMERSKI S. Individuals on alert: digital epidemiology and the individualization of surveillance[J]. Life Science, Society and Policy, 2018, 14(1):13-26.

[39] DENECKE K. An ethical assessment model for digital disease detection technologies[J]. Life Sciences, Society and Policy, 2017, 13(16):1-11.

[40] 苏玉娟. 政府数据治理的五重系统特性探讨[J]. 理论探索, 2016(2):71-75.

[41]郭蓉. 从技术理性到行政伦理——大数据时代智慧治理的伦理反思[J]. 道德与文明, 2018(6):19-25.

[42]张郁安,杨筱敏. 疫情下的数据利用和个人信息保护再权衡[N]. 人民邮电报, 2020-02-17.

[43]DAVID J. Hand aspects of data ethics in a changing world: where are we now? [J] Big Data, 2018(2):176-190.

[44]雷瑞鹏,邱仁宗. 数据共享是道德律令[J]. 自然辩证法研究, 2018 (1): 46.

[45]李伦,黄关. 数据主义与人本主义数据伦理[J]. 伦理学研究, 2019(2):102-107.

[46]刘士国,熊静文. 健康医疗大数据中隐私利益的群体维度[J]. 法学论坛, 2019, 34(3):125-135.

[47]王丽莎. 互联网医疗大数据的法律与伦理规制研究[J]. 中国医学伦理学, 2017, 20(11):1322-1325, 1342.

[48]约叔华·A. 克鲁尔,乔安娜·休伊,索伦·巴洛卡斯,等. 可问责的算法[J]. 地方立法研究, 2019(4):102-150.

[49]徐凤. 人工智能算法黑箱的法律规制——以智能投顾为例展开[J]. 东方法学, 2019(6):78-86.

[50]沈伟伟. 算法透明原则的迷思——算法规制理论的批判[J]. 环球法律评论, 2019(6).

[51] NEW J, CASTRO D. How policymakers can foster algorithmic accountability[OL/EB]. [2020-03-25]. <http://www2.datainnovation.org/2018-algorithmic-accountability.pdf>.

[52] THALEN M. AI beat the CDC to the punch on coronavirus warnings[EB/OL].  
[2020-02-20].<https://www.dailydot.com/debug/blue-dot-ai-beat-cdc-coronavirus-warnings/>.

[53] GAGLIARDO C. Digital epidemiology and harnessing big data for infectious diseases and public health[EB/OL].

[2020-04-11].<https://www.pids.org/news/451-digital-epidemiology.html>.

[54]冯登国, 张敏, 李昊. 大数据安全与隐私保护[J]. 计算机学报, 2014, 37(1):246-258.

**作者简介:** 丁波涛, 1977 年生, 博士, 副研究员。研究方向: 信息资源管理, 智慧城市, 智慧治理。

**录用日期:** 2020-12-02

